

SYSTEM AND PROCESS FOR PAN-GENOMIC  
DETERMINATION OF MACROMOLECULAR ATOMIC STRUCTURE

ABSTRACT OF THE DISCLOSURE

5 A process for pan-genomic determination of three-dimensional  
macromolecular atomic structures uses a unique combination of  
components. All known structural information, sequence  
information and functional information are systematically  
10 organized into a genomics database. Advanced tools of  
bioinformatics are used to cluster all known gene products  
into families of homologous sequences. Simultaneously, in  
parallel for each such family, a few cDNAs from appropriately  
representatives species are cloned into expression vectors for  
a few expressions systems. Constructs are then screened for  
15 expression, and those that are effective advance to the  
preparative step. Expressed proteins are prepared, purified  
and characterized. Purified proteins are set to crystallize  
in parallel against crystallization screens. Crystals that  
grow are tested for suitable diffraction characteristics. A  
20 suitable crystal is frozen, and diffraction data are measured  
using the multiwavelength anomalous diffraction (MAD) method  
at a synchrotron which uses undulator beamlines for high-  
throughput crystallography. Diffraction data are analyzed by  
the MAD phasing method, an atomic model is built, and the  
25 model is refined against the diffraction data. The refined  
model is analyzed in the context of (1) sequence information  
from other family members, (2) all other known 3D structures,  
and (3) functional motifs. It is also analyzed for surface  
characteristics with the aim to define active sites and  
30 macromolecular contact sites. For relevant structures, the  
active site properties are used to define classes of compounds  
predicted to have binding potency. Computational tools for  
homology model building are used to develop models for  
homologs. ~~The homology models may be used in target~~  
35 selection, drug design, or design of more appropriate  
constructs for experimental analysis. The ensemble of all  
known structures is used to further advance the effectiveness  
of the bioinformatics tools.